

УДК 519.24

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ СО СТОХАСТИЧЕСКИМ ТРЕНДОМ

В. М. Андриенко

Кандидат экономических наук*

Контактный тел.: (048) 64-13-29, 067-488-52-15**E-mail:** andrienko.v@gmail.com**Е. А. Арсирый**

Кандидат технических наук*

E-mail: arsiriy@te.net.ua*Кафедра экономической кибернетики
и информационных технологийНациональный политехнический университет
пр. Шевченко, 1, г. Одесса, Украина, 65044

Робота присвячена аналізу і моделюванню часових рядів із стохастичним трендом. Розглядаються проблеми виявлення тренда і ідентифікації моделі. Наводяться результати моделювання реальних рядів, отримані на математичній і нейромережовій моделях.

Ключові слова: часовий ряд, стохастичний тренд, нейромережовий аналіз.

Работа посвящена анализу и моделированию временных рядов со стохастическим трендом. Рассматриваются проблемы обнаружения тренда и идентификации модели. Приводятся результаты моделирования реальных рядов, полученные на математической и нейросетевой моделях.

Ключевые слова: временной ряд, стохастический тренд, нейросетевой анализ.

Work is devoted an analysis and design of temporal rows with stochastic trend. The problems of finding out trena and authentication of model are examined. Results over of design of the real rows got on mathematical and neuron models are brought.

Keywords: temporal row, stochastic trend, analysis of neuron network.

1. Введение

Большинство математико-статистических методов имеет дело с моделями, в которых наблюдения предполагаются независимыми и одинаково распределенными. При этом основное внимание уделяется проблемам идентификации моделей, отбору эндогенных и экзогенных показателей, но почти не обращается внимания на формальный анализ структуры исходных статистических рядов. Зависимость между наблюдениями чаще всего рассматривается как помеха в эффективном применении этих методов. Однако разнообразные данные в экономике, социологии, финансах, коммерции и других сферах человеческой деятельности поступают в форме *временных рядов*, в которых наблюдения взаимно зависимы, и характер этой зависимости как раз и представляет главный интерес для исследователя.

Временным рядом называется совокупность наблюдений показателя в различные моменты времени. Обычно временной ряд рассматривают как выборку из последовательности случайных величин X_t , где t принимает целочисленные значения от 1 до T . Совокупность случайных величин $\{X_t, t \in [1, T]\}$ называют *дискретным случайным процессом* или *стохастическим процессом* [1]. Принципиальные отличия временного ряда от последовательности наблюдений, образующих случайную выборку, заключаются в следующем:

- члены временного ряда не являются независимыми;
- члены временного ряда не обязательно являются одинаково распределенными.

Это означает, что свойства и правила статистического анализа случайной выборки нельзя распространять на временные ряды. С другой стороны, взаимозависимость членов временного ряда создает свою специфическую базу для построения прогнозных значений анализируемого показателя по эмпирическим значениям.

Аналитически временной ряд можно выразить уравнением вида:

$$X(t) = f(t) + S(t) + \varepsilon(t), \quad (1)$$

где $f(t)$ — тренд (долговременная тенденция) развития; $S(t)$ — сезонная (периодическая) компонента; $\varepsilon(t)$ — случайная величина (случайная компонента).

Функция $f(t)$ определяет общую тенденцию развития изучаемого явления. Тренд может быть выражен как детерминированной, так и случайной функциями, либо их комбинацией. Компоненты временного ряда $f(t)$, $S(t)$ и $\varepsilon(t)$ ненаблюдаемы. Они являются теоретическими величинами. Выявление этих компонент и является задачей анализа.

При построении моделей связей необходимо решать вопрос об отнесении каждого из рассматриваемых рядов к классу рядов, стационарных относительно детерминированного тренда, или к классу рядов, имеющих стохастический тренд. Принципиальное различие между этими двумя классами выражается в том, что в первом случае вычитание из ряда соответствующего детерминированного тренда приводит к стационарному ряду, тогда как во втором — вычитание детерминированной составляющей оставляет ряд нестационарным из-за наличия у него стохастического тренда.

Стохастические тренды характерны для макроэкономических показателей описывающих поведение того или иного бизнес процесса во времени, показателей фондового рынка, денежных потоков, изменений ежедневных остатков на складах, в инструментальных кладовых, магазинах. Для анализа бизнес процессов, представленных временными рядами со стохастическим трендом с успехом используются методы и модели интеллектуального анализа данных (статистические методы и методы машинного обучения), при этом предлагается не математический аппарат, а последовательность действий, которую необходимо выполнить исследователю для построения прогнозной модели (извлечения знаний из временного ряда).

2. Методы анализа трендовых рядов

В рядах с детерминированным трендом влияние предыдущих шоковых воздействий (резких изменений) затухает с течением времени, а в рядах со стохастическим трендом такое затухание отсутствует, и каждый отдельный шок влияет с одинаковой силой на все последующие значения ряда. Детерминированные компоненты $f(t)$ и $S(t)$ обнаруживаются с помощью спектрального анализа. Детерминированная составляющая имеет неограниченную спектральную плотность на низких частотах, а при наличии периодических составляющих спектральная плотность имеет максимумы (пики). Бокс и Дженкинс [2] предложили компоненту $f(t)$ исключать дифференцированием, то есть вычислением последовательных разностей. Если ряд после вычисления d последовательных разностей приводится к стационарному, то такой ряд называют ARIMA(p, d, q) (авторегрессионные проинтегрированные скользящего среднего — Autoregressive Integrated Moving Average) порядка (p, d, q), которые моделируют различные ситуации, встречающиеся при анализе стационарных и нестационарных рядов. При этом p параметр AR — части, d — степень интеграции, q — параметр MA — части. Периодическую составляющую удаляют вычитанием ее из ряда.

Стохастический тренд обнаруживается с помощью спектрального и автокорреляционного анализа. Автокорреляционная функция в этом случае медленно убывает, а периодограмма на низких частотах неограниченно возрастает. Такие ряды называют «временными рядами с долговременной корреляционной зависимостью (time series with long memory)». В работах зарубежных ученых, в первую очередь, С. W. Granger, J. R. Hosking, P. M. Robinson, R. Beran [3], был предложен новый класс моделей ARFIMA(p, d, q), допускающий возможность нецелого параметра d и получивший название авторегрессионный дробно-интегрированный процесс скользящего среднего. Характеристики таких временных рядов обладают важными свойствами, например, X_t является стационарным и обратимым для $d \in (-1/2, 1/2)$. При этом X_t можно представить в виде [4]:

$$\Phi(B)(1-B)^d X_t = \Theta(B)\varepsilon_t, \quad (2)$$

где $B = X_{t-1}/X_t$ — оператор сдвига назад,

$$\Phi(B) = 1 - \sum_{j=1}^p \phi_j B^j; \quad \Theta(B) = 1 + \sum_{j=1}^q \theta_j B^j$$

— полиномы, предполагается, что все корни уравнений $\Phi(z) = 0$, $\Theta(z) = 0$ по модулю больше единицы,

$$(1-B)^d = 1 - dB + \frac{d(d-1)}{2} B^2 - \dots = \sum_{j=0}^{\infty} \psi_j B^j,$$

$$\psi_j = \prod_{0 \leq k < j} \frac{k-1-d}{k} = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}, \quad j=0, 1, 2, \dots,$$

ε_t — гауссовский белый шум $WN(0, \sigma^2)$, $\Gamma(x)$ — гамма-функция, которая определяется формулой

$$\Gamma(x) = \begin{cases} \int_0^{\infty} t^{x-1} e^{-t} dt, & x > 0, \\ x^{-1} \Gamma(1+x), & x < 0, x \neq -1, -2, \dots \end{cases} \quad (3)$$

В частном случае, при $p, q = 0$ и $d \in (-1/2, 1/2)$, процесс имеет вид:

$$X_t = (1-B)^{-d} \varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad (4)$$

где $\psi_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}$.

Оценку параметра d можно получить из равенства

$$d = H - 0,5, \quad (5)$$

где H — показатель Херста [5].

Если $H > 0,5$, то это также указывает на наличие стохастического тренда, который не может быть удален дифференцированием. Для вычисления показателя H известный британский гидролог Х. Е. Херст предложил метод нормированного размаха (R/S-анализ). Кроме того, на основе R/S-анализа вычисляется R/S-статистика для проверки статистической гипотезы об отсутствии долгосрочной зависимости. Процедура проверки гипотезы подробно описана в [6]. R/S-анализ является простым процессом, но он требует обработки большого количества данных. Благодаря развитию компьютерных технологий, созданы программные продукты, с помощью которых вычисляют коэффициент Херста. Один из них Fractan 4.4. Продукт распространяется бесплатно и помещен на многих Internet-сайтах, например [7]. Для осуществления моделирования с помощью Fractan потребуется более трех тысяч эмпирических данных, что не всегда доступно исследователю.

В условиях ограниченных реальных данных и благодаря простоте практического применения, в настоящее время особую популярность для анализа и прогнозирования временных рядов приобретают коммерческие (платные) и некоммерческие (бесплатные) приложения — нейроимитаторы, использующие для анализа и прогноза методы машинного обучения на основе нейронных сетей различных архитектур. Среди коммерческих нейроимитаторов особенно популярны NeuroSolutions — фирмы NeuroDimension, Inc., NeuralWorks Professional U/Plus с модулем UDND — фирмы Neural Ware, Inc., NeuroShell2 — фирмы Ward Systems Group, Statistica Neural Networks — фирмы StatSoft Russia. Исследователю доступны также свободно распространяемые нейроимитаторы: Excel Neural Package (trial-версия), расширяющий функциональные возможности работы с данными Microsoft Excel 97, Neural Network Wizard — программная реализация многослойной нейронной сети обратного распространения,

Neuro Laboratory, Neuro Office'98, NeuroPro версий 0.1 и 0.25. Популярность у исследователей приобретают инструментальные средства библиотек Artificial neural network (ANN) toolbox SciLab и Neural Network (NN) Toolbox Matlab.

Однако, применение нейроимитаторов для решения практических задач является «скорее искусством, чем наукой» [8], так как выбор многочисленных параметров, которые требуются для моделирования, осуществляется только на основе личного практического опыта исследователя. Тем не менее, на практике для моделирования и прогнозирования предпочитают использование именно нейроимитаторов.

3. Методика исследования

Исследуем динамику цен акций на момент закрытия торгов, в группе ММВБ нескольких российских компаний за 2000–2010 годы (данные сайта <http://stocks.investfunds.ru/>). Показатель Херста, который был определен по этим данным оказался для всех компаний больше 0,5 (табл. 1), то есть анализируемые временные ряды являются временными рядами с долговременной памятью.

Таблица 1

Значения показателя Херста

Наименование компании	Аэрофлот	Газпром	Лукойл	Камаз
Показатель Херста	0,64	0,71	0,58	0,68

О наличии стохастического тренда свидетельствуют и оценки спектральной плотности и корреляционной функции: периодограмма и коррелограмма (рис. 1 и рис. 2). Коррелограмма медленно убывает на лагах 1–20, периодограмма неограниченно возрастает на низких частотах.

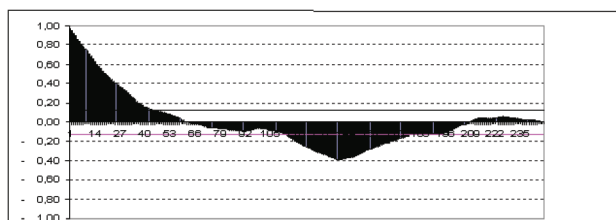


Рис. 1. Коррелограмма значений цен акций компании «Газпром»

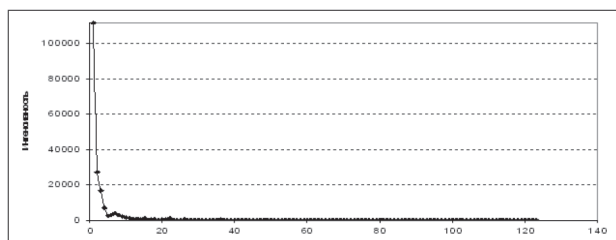


Рис. 2. Периодограмма значений цен акций компании «Газпром»

Аналогичный результат получен и для других рассматриваемых компаний.

Для моделирования применим процесс (4). Для этого обозначим P_i ($i=1,2,...,N$) — ряд котировок ценной бумаги. Образует из данного ряда последовательность

$$R_i = \ln \left(\frac{P_i}{P_{i-1}} \right). \quad (6)$$

Цены закрытия P_i вычисляются на следующем торговом интервале по формуле:

$$P_i = P_{i-1} \exp(R_i). \quad (7)$$

Выражение для R_i разделим на две части: среднее значение \bar{R}_i и случайную составляющую ε_i :

$$R_i = \bar{R} + \varepsilon_i, \quad (8)$$

где ε_i случайный процесс (4),

$$\varepsilon_i = \sum_{k=0}^{\infty} \psi_k \xi_{i-k}, \quad \psi_k = \frac{\Gamma(k+H-0,5)}{\Gamma(k+1)\Gamma(H-0,5)}, \quad k \geq 1, \quad (9)$$

в котором ξ_i — случайная величина, распределенная по нормальному закону с нулевым средним значением и среднеквадратическим отклонением σ , $\psi_0=1$, H — показатель Херста, $\Gamma(x)$ — гамма-функция, которая определяется формулой (3). По существу, формула (9) представляет собой свертку процесса Гаусса с бесконечным рядом, то есть бесконечное скользящее среднее процесса Гаусса, моделирующее процесс с памятью ARFIMA(0,d,0), где $d=H-0,5$, $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$.

Результаты моделирования помещены ниже, в табл. 2. В качестве характеристики точности прогноза принята *средняя ошибка аппроксимации*, которая выражается в процентах относительно фактических значений признака, и определяется по формуле вида:

$$\bar{\delta}^* = \frac{1}{n} \sum_{t=1}^n \frac{|P_t - P_t^*|}{P_t} \cdot 100\%, \quad (10)$$

P_t^* — прогнозные значения, P_t — фактические значения исследуемого признака.

Данный показатель является относительным показателем точности прогноза и не отражает размерность изучаемых признаков, выражается в процентах и на практике используется для сравнения точности прогнозов полученных как по различным моделям, так и по различным объектам.

Теперь проведем машинное обучение нейронных сетей на тех же данных временного ряда. В данном случае целесообразно применить нейронную сеть Элмана, которая реализована в NNToolbox Matlab. Эта сеть, состоящая из двух слоев, в которой скрытый слой содержит рекуррентные связи со слоем контекстных единичных задержек. Эти контекстные элементы сохраняют выходы скрытых нейронов на один шаг во времени, позволяя им помнить свои предыдущие действия и реализовывать задачи обучения, которые разворачиваются во времени. Подобный учет предыстории наблюдаемых процессов актуален для моделирования временных рядов с памятью. Сеть Элмана с одним нейроном в скрытом слое и одним контекстным нейроном создана с помощью функции:

`net = newelm([1 20],[1 1],{'tansig','tansig'},'trainlm').`

В качестве функции активации использовалась функция гиперболического тангенса — `tansig`. Для обучения

Таблица 2

Результаты моделирования

Аэрофлот			Газпром			Лукойл		
Цена акций (руб.)			Цена акций (руб.)			Цена акций (руб.)		
Фактическая	Модель		Фактическая	Модель		Фактическая	Модель	
	Математическая	Нейросетевая		Математическая	Нейросетевая		Математическая	Нейросетевая
88,49	89,73	89,93	307,00	303,87	302,15	1 666,00	1719,42	1717,71
86,00	89,70	90,04	296,00	304,33	312,74	1 600,00	1720,04	1718,74
90,79	89,63	90,23	307,20	302,28	310,76	1 736,99	1721,37	1717,72
92,70	89,48	90,38	316,23	305,87	320,50	1 818,00	1723,81	1715,74
90,00	89,43	90,46	305,32	302,99	323,43	1 737,00	1724,08	1714,54
90,40	89,33	90,80	312,83	305,04	328,54	1 775,00	1726,18	1715,03
90,87	89,26	90,88	305,95	305,24	330,11	1 736,00	1726,55	1715,81
90,00	89,27	90,88	290,95	303,81	327,82	1 668,00	1728,85	1714,14
93,47	89,27	91,05	305,20	303,59	326,50	1 739,99	1729,58	1712,97
101,49	89,21	91,07	312,88	304,16	337,22	1 768,00	1730,52	1714,64
98,00	89,10	91,20	302,49	300,86	339,95	1 725,02	1732,17	1714,60
99,40	89,06	91,42	297,70	301,33	333,22	1 715,01	1733,10	1717,52
98,86	88,92	91,52	289,09	300,96	330,82	1 654,00	1735,40	1717,66
98,58	88,97	91,59	286,31	301,02	331,31	1 638,00	1735,12	1714,31
99,71	88,84	91,65	292,25	302,40	340,75	1 697,00	1737,01	1714,06
103,50	88,74	91,88	302,82	301,89	334,23	1 769,00	1738,63	1712,81
106,20	88,63	91,95	307,81	303,11	329,48	1 807,04	1739,88	1714,12
Средняя ошибка $\bar{\delta}^*$, %	6,61	5,09	Средняя ошибка $\bar{\delta}^*$, %	2,15	4,12	Средняя ошибка $\bar{\delta}^*$, %	2,78	2,88

применен алгоритм Левенберга — Марквардта — trainlm, основанный на оценке матрицы Гессе и относящийся к методам обучения второго порядка. Результаты моделирования представлены ниже, на рис. 3 и в табл. 2. Из таблицы и графиков рисунка видно, что модель полностью отражает тенденцию данных.

Средняя относительная ошибка аппроксимации для всех компаний меньше 10 %, что свидетельствует о высокой оценке точности прогноза. Для компании «Газпром» результат нейросетевого моделирования немного хуже, чем на математической модели, но зато не потребовалось большого числа эмпирических данных и не пришлось выполнять трудоемких вычислений.

В данной работе для автокорреляционного и спектрального анализа эмпирических данных использовался компьютерный пакет AtteStat, который является надстройкой приложения Microsoft Excel. После установки он загружается автоматически вместе с Microsoft Excel. Пакет распространяется свободно. В настоящее время

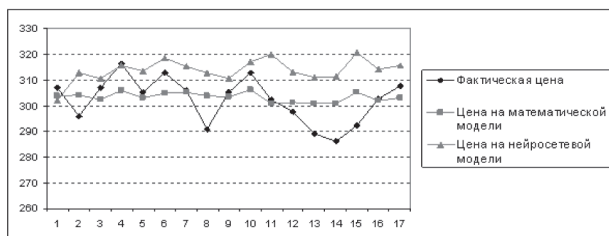


Рис. 3. Фактические и прогнозные значения цены акций компании «Газпром»

популярным является пакет EvIEWS. В нем заложены широкие возможности при анализе данных, представленных в виде временных рядов. Программа EvIEWS имеет удобный и дружелюбный интерфейс, проста в обращении и интерпретации результатов. Она дает возможность осуществлять импорт/экспорт из файлов Microsoft Excel и производить априорные оценки адекватности модели, используя различные информационные критерии. Информацию о пакете можно найти на сайте [9].

4. Выводы

Анализ временных рядов с долговременной памятью представляет не только теоретический интерес. Он дает возможность исследовать устойчивость показателя, поскольку исследование устойчивости сводится к определению типа временного ряда эмпирических значений показателя:

- ряд стационарный относительно детерминированного тренда, что свидетельствует о стабильности исследуемого показателя;
- ряд содержит стохастический тренд, что является основанием для противоположного заключения.

Для построения адекватной модели, которую можно использовать для описания динамики временного ряда и прогнозирования его будущих значений, необходимо выяснить природу этого ряда. Моделирование и прогнозирование на основе нейроимитатора дает достаточно высокий уровень точности прогнозов.

Литература

1. Канторович Г. Г. Анализ временных рядов [Текст] / Г. Г. Канторович // Экономический журнал ВШЭ. — № 2. — 2002. — С. 252—273.
2. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление [Текст] / Дж. Бокс, Дженкинс Г. — Вып. 1, 2. — М.: Мир, 1974. — 197 с.
3. Granger C. W. J. Some Properties of Time Series Data and Their Use in Econometric Model Specification [Текст] / C. W. J. Granger // Journal of Econometrics, 1981. — Vol. 16. — № 1. — P. 121—130.
4. Леоненко М. М. Теоретико-ймовірнісні та статистичні методи в економетриці та фінансовій математиці [Текст] / М. М. Леоненко, Ю. С. Мішура, В. М. Пархоменко, М. Й. Ядренко. — К.: Інформтехніка, 1995. — 380 с.
5. Петерс Э. Фрактальный анализ финансовых рисков. [Текст] / Э. Петерс. — М.: Интернет-трейдинг, 2004. — 304 с.
6. Lo A. W. Long Term Memory in Stock Market Prices [Текст] / A. W. Lo // Econometrica. — 1991. — № 59. — P. 1279—1313.
7. Softodrom.ru — Лучшие программы для Windows, Windows Mobile, Linux и других ОС [Электронный ресурс] [Текст] / Режим доступа: <http://soft.softodrom.ru/ap/Fractal-p44195>. — 10.05.2011.
8. Хайкин С. Нейронные сети: полный курс [Текст]: пер. с англ. / под ред. д. т. н. Н. Н. Кузусль. — 2-е изд., испр. — М.: ООО «И. Д. Вильямс», 2006. — 1104 с.
9. Exnews. Екцентричні новини — Програми. Софт [Електронний ресурс] [Текст] / Режим доступа: <http://exnews.org/grabsoft/1146907935-views-enterprise-edition>. — 10.05.2011.

УДК 621.371.3

СОПОСТАВЛЕНИЕ ЭКСПЕРИМЕНТАЛЬНЫХ И РАСЧЕТНЫХ МЕТЕОРНЫХ АМПЛИТУДНО-ВРЕМЕННЫХ ХАРАКТЕРИСТИК

И. Е. Антипов

Доктор технических наук, профессор, заведующий кафедрой*

E-mail: i_ant@mail.ru**Е. Ю. Бондарь**

Кандидат технических наук, младший научный сотрудник

Научно-учебный центр*

E-mail: gxaski@mail.ru**О. А. Соляник**

Кандидат технических наук, старший научный сотрудник

Научно-учебный центр*

E-mail: secort@kture.kharkov.ua**Набиль Оде Сорох**

Аспирант*

E-mail: bassra55@yahoo.com

*Кафедра основ радиотехники

Харьковский национальный университет радиоэлектроники

пр. Ленина, 14, г. Харьков, Украина, 61166

Контактный тел.: (057) 700-22-84

Розглянуто засіб зіставлення результатів моделювання метеорних АЧХ з експериментальними даними.

Ключові слова: метеорна АЧХ, зони Френеля, лінійна регресія.

Рассмотрен способ сопоставления результатов моделирования метеорных АВХ с экспериментальными данными.

Ключевые слова: метеорная АВХ, зоны Френеля, линейная регрессия.

A method for comparing the data of experimental and estimated meteor ATC.

Keywords: meteor ATC, Fresnel zone, linear regression.

1. Введение

Для исследования метеорных явлений применяется радиолокационный метод, поскольку метровые радио-

волны отражаются от ионизированного следа, образуемого при сгорании в атмосфере метеорной частицы. Поведение амплитуды принимаемого сигнала описывается амплитудно-временной характеристикой (АВХ).